HOW BIG ARE THE MORTALITY REDUCTIONS PRODUCED BY CANCER SCREENING?

WHY DO SO MANY TRIALS SAY 20%?

James A. Hanley

Department of Epidemiology, Biostatistics & Occupational Health, McGill University

Symposium: Randomized Trials of Cancer Screening: How Useful are They?
3rd North American Congress of Epidemiology, June 23, 2011

Summary: the 3 points I wish to make

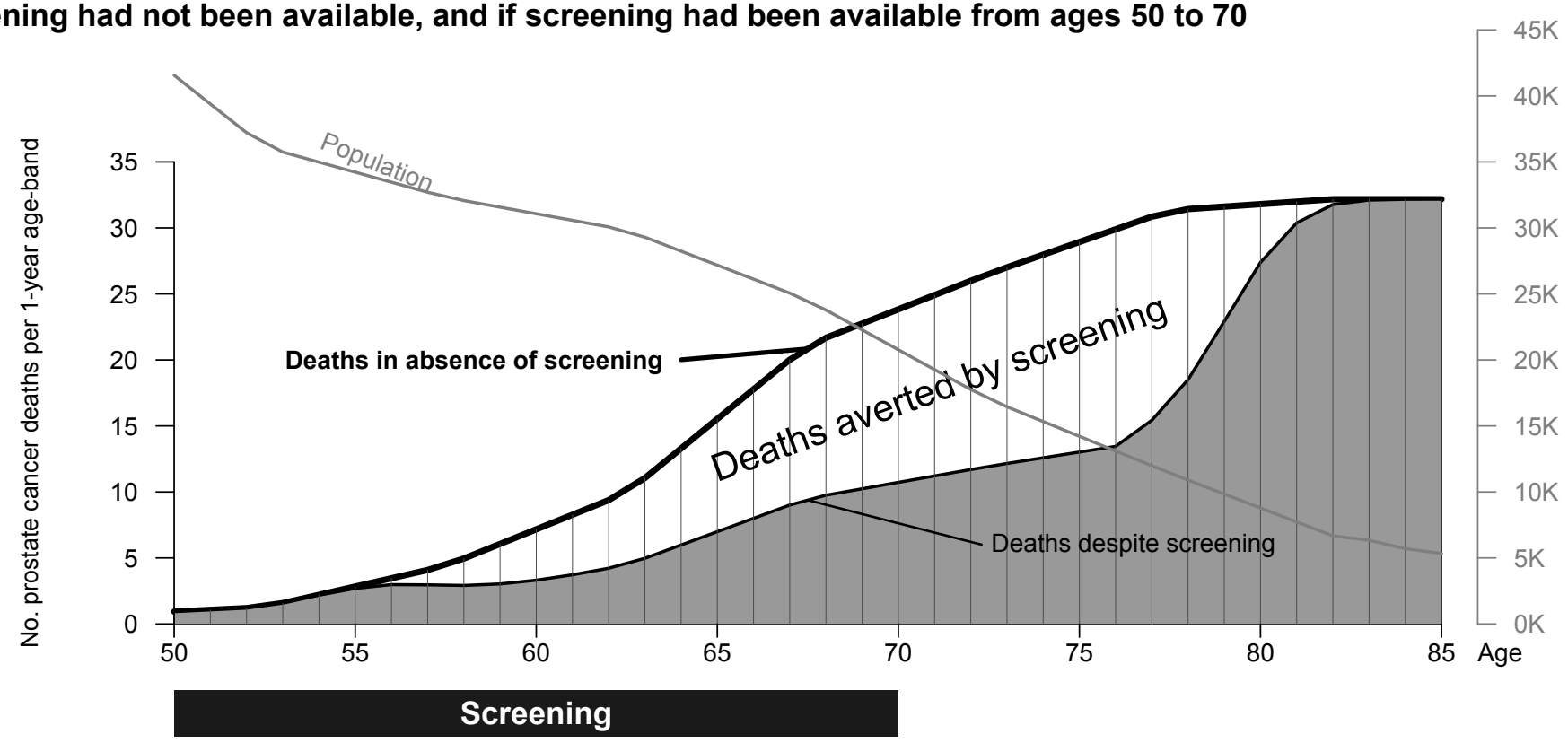
- With their blindness to the delay until the reductions in mortality are expressed, the prevailing design and data-analysis of cancer screening trials ***under-estimate*** the mortality reductions that ***would be produced by a sustained screening program***
- P-value-driven stopping rules exacerbate the underestimation
- We *might* be able to avoid such misleading numbers if we
  (i) recognize the issue, and avoid the standard RCT paradigm
  (ii) run trials with sufficient rounds of screening and sufficient follow-up
  (iii) spend major portion of career waiting to measure real reductions
  (iv) analyze the data using time-specificity
  (v) focus on the parameters that describe impact of 1 round of screening

Outline

- The mortality reductions produced by a screening regimen: what payers want to know
- European Randomized Study of Screening for Prostate Cancer
- Data-analysis practice in other cancer screening trials
- How to stop a screening RCT at a 20% mortality reduction? [Theorem]
- A way ahead?

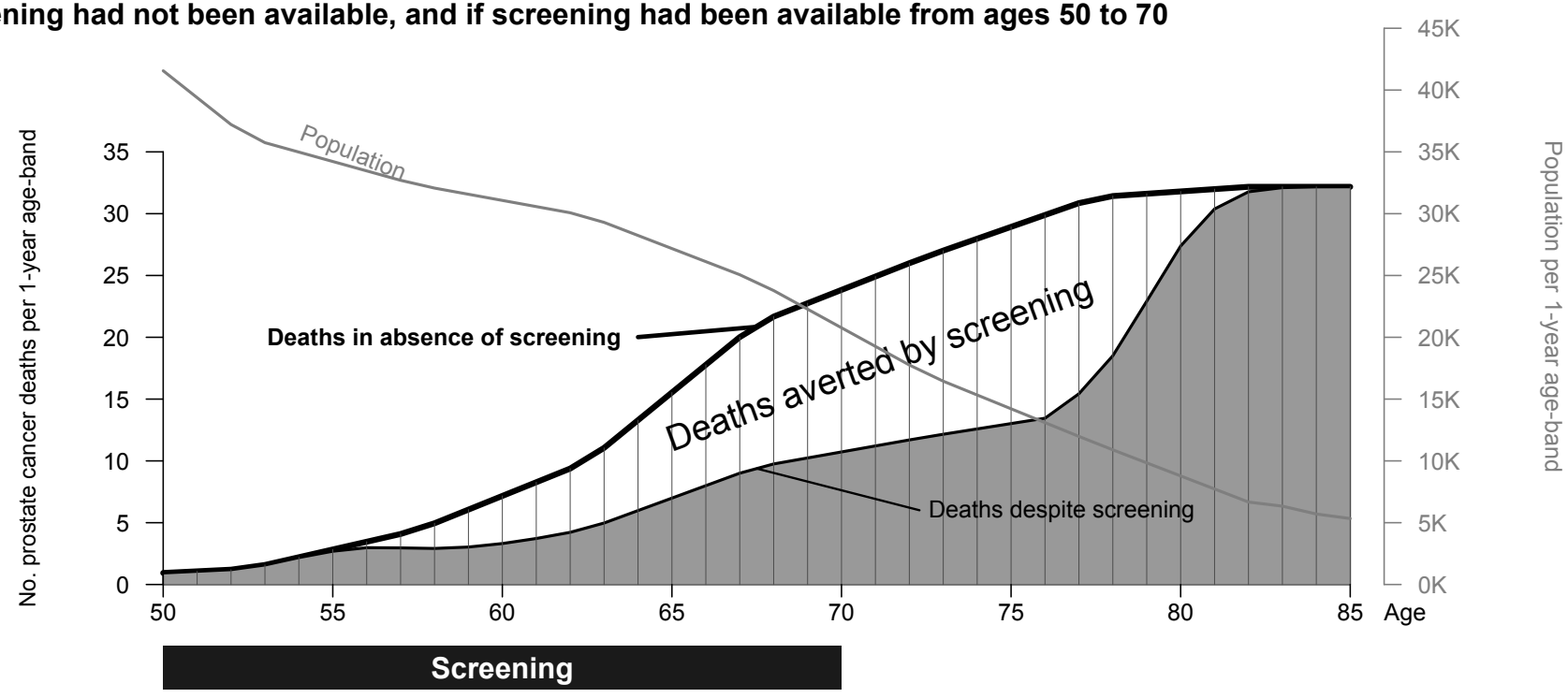
What payers would like to know...

(a) Age-specific numbers of prostate cancer deaths in a steady state population with a given age-structure, if screening had not been available, and if screening had been available from ages 50 to 70

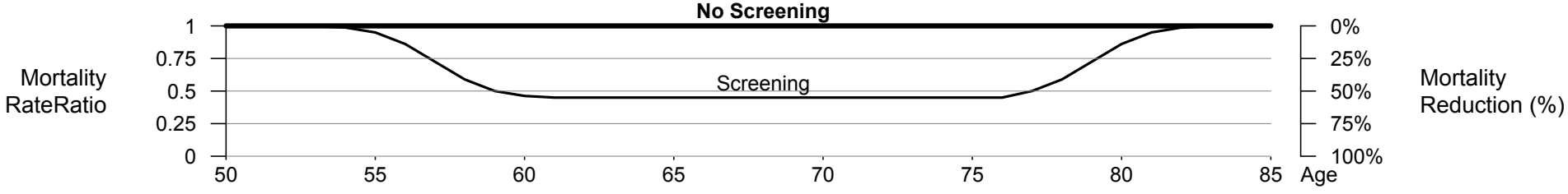


They could arrive at these numbers if they had...

(a) Age-specific numbers of prostate cancer deaths in a steady state population with a given age-structure, if screening had not been available, and if screening had been available from ages 50 to 70



(b) The corresponding age-specific prostate cancer mortality rate ratios



Can they obtain these (or asymptote) from published reports?

ORIGINAL ARTICLE

Screening and Prostate-Cancer Mortality in a Randomized European Study

Fritz H. Schröder, M.D., Jonas Hugosson, M.D., Monique J. Roobol, Ph.D., Teuvo L.J. Tammela, M.D., Stefano Ciatto, M.D., Vera Nelen, M.D., Maciej Kwiatkowski, M.D., Marcos Lujan, M.D., Hans Lilja, M.D., Marco Zappa, Ph.D., Louis J. Denis, M.D., Franz Recker, M.D., Antonio Berenguer, M.D., Liisa Määttänen, Ph.D., Chris H. Bangma, M.D., Gunnar Aus, M.D., Arnauld Villers, M.D., Xavier Rebillard, M.D., Theodorus van der Kwast, M.D., Bert G. Blijenberg, Ph.D., Sue M. Moss, Ph.D., Harry J. de Koning, M.D., and Anssi Auvinen, M.D., for the ERSPC Investigators*

ABSTRACT

BACKGROUND

The European Randomized Study of Screening for Prostate Cancer was initiated in the early 1990s to evaluate the effect of screening with prostate-specific–antigen (PSA) testing on death rates from prostate cancer.

METHODS

We identified 182,000 men between the ages of 50 and 74 years through registries in seven European countries for inclusion in our study. The men were randomly assigned to a group that was offered PSA screening at an average of once every 4 years or to a control group that did not receive such screening. The predefined core age group for this study included 162,243 men between the ages of 55 and 69 years. The primary outcome was the rate of death from prostate cancer. Mortality follow-up was identical for the two study groups and ended on December 31, 2006.

The authors' affiliations are listed in the Appendix. Address reprint requests to Dr. Schröder at the Erasmus Medical Center, P.O. Box 2040, Rotterdam 3000 CA, the Netherlands, or at secr.schroder@erasmusmc.nl.

*Members of the European Randomized Study of Screening for Prostate Cancer (ERSPC) are listed in the Appendix.

This article (10.1056/NEJMoa0810084) was published at NEJM.org on March 18, 2009.

N Engl J Med 2009;360:1320-8.

...

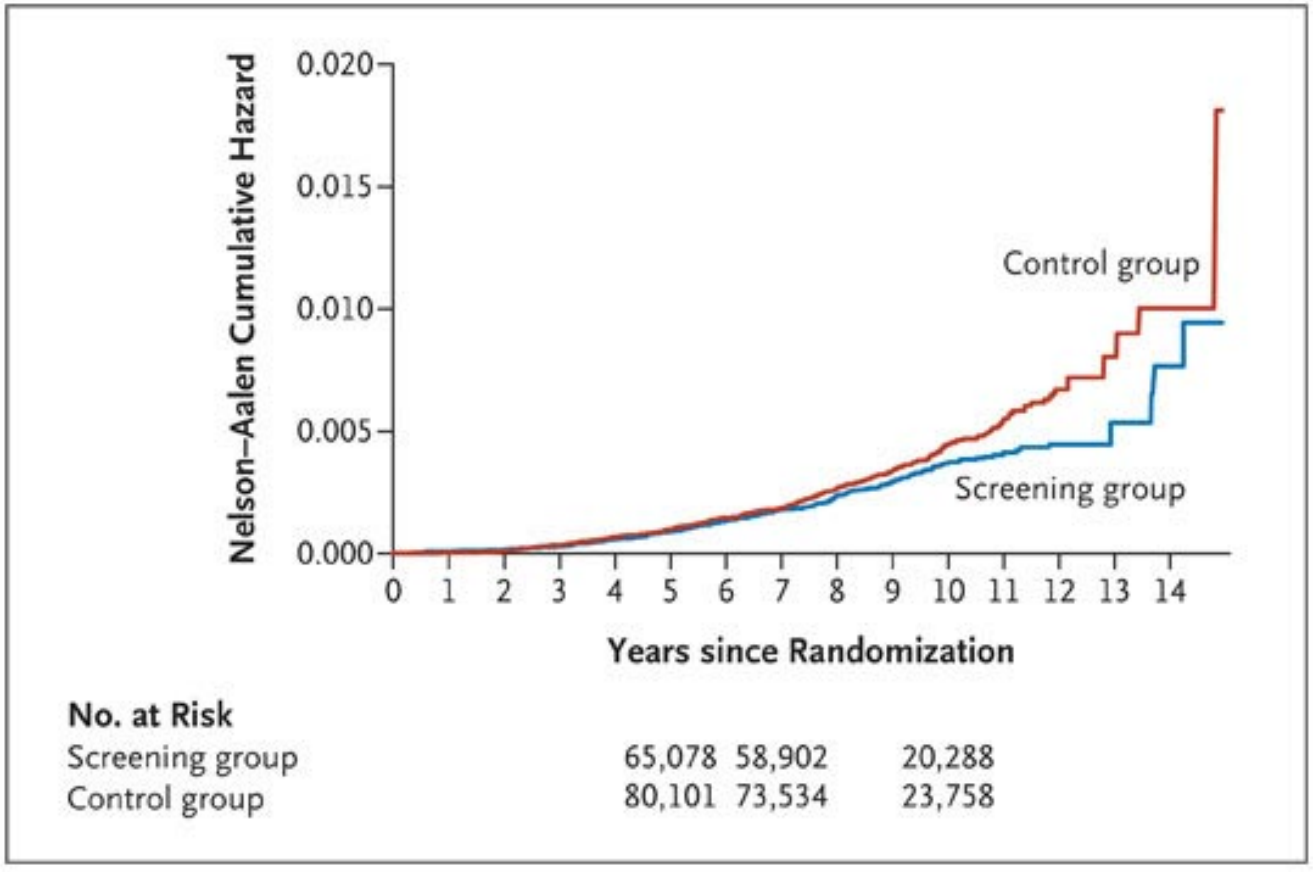
RESULTS

"During a median follow-up of 9 years, the prostate cancer mortality rate ratio in the screening group, as compared with the control group, was 0.80 (95% confidence interval [CI], 0.65 to 0.98; adjusted P=0.04). (...) "

CONCLUSIONS

"PSA-based screening reduced the rate of death from prostate cancer by 20%. (...) "

Cumulative Risk of Death from Prostate Cancer.



As of December 31, 2006, with an average follow-up time of 8.8 years, there were 214 prostate-cancer deaths in the screening group and 326 in the control group. (...) The **adjusted rate ratio** for death from prostate cancer in the screening group was **0.80** (95% CI, 0.65 to 0.98; P=0.04).

Cumulative vs. Year-specific Mortality...

in 100,000 men

(average age at entry: 62 years)

if screened using PSA test

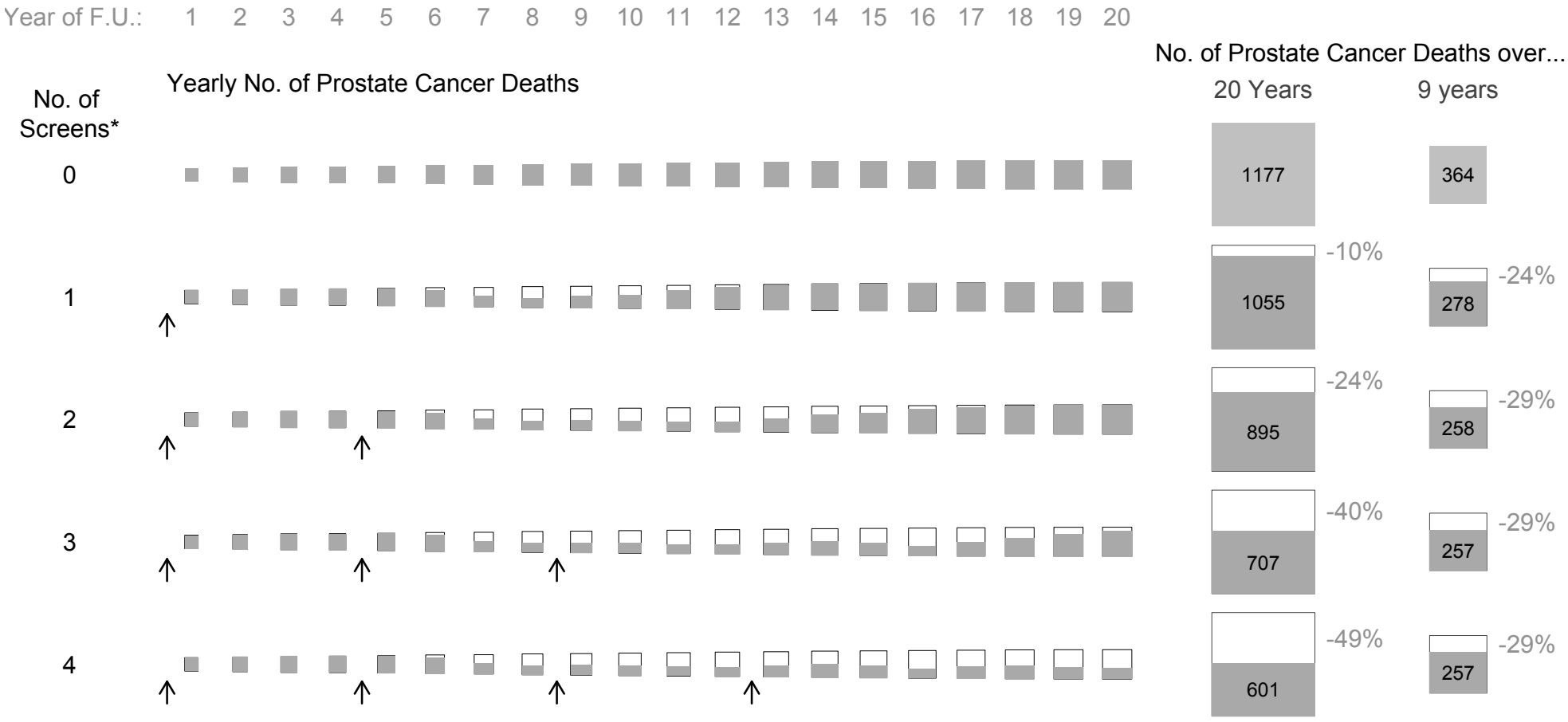
0, 1, 2, 3, or 4 times,

tests 4 years apart

and followed for (9) 20 years

HYPOTHETICAL DATA

Cumulative & Year-specific results, if screen 0,1,...,4 times, q 4y [HYPOTHETICAL]



* Each arrow indicates the timing of a screen for prostate cancer.

(B) Year-specific Rate Ratios & Percent Reductions [HYPOTHETICAL]

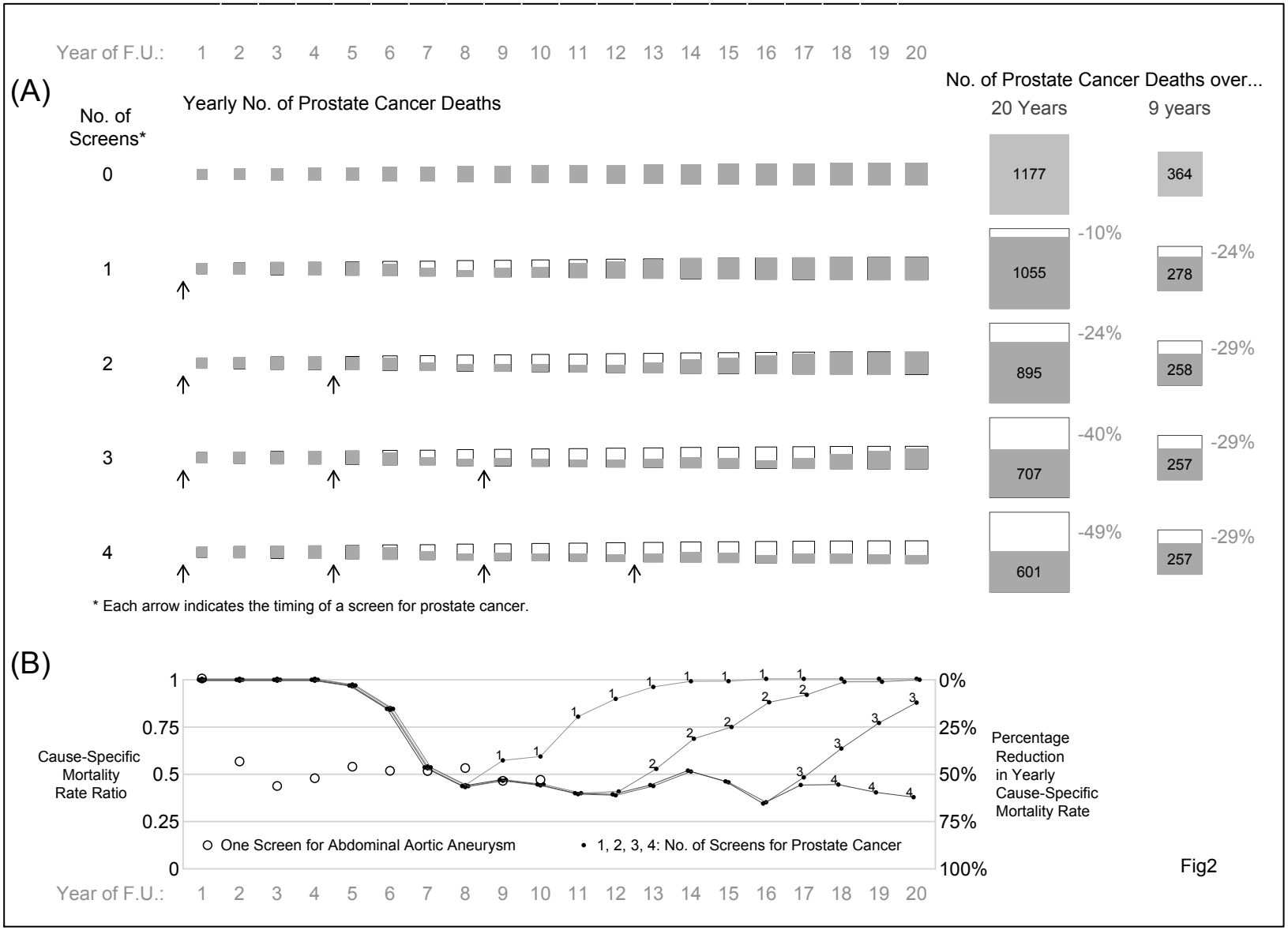


Fig2

RE-ANALYSIS OF ERSPC DATA

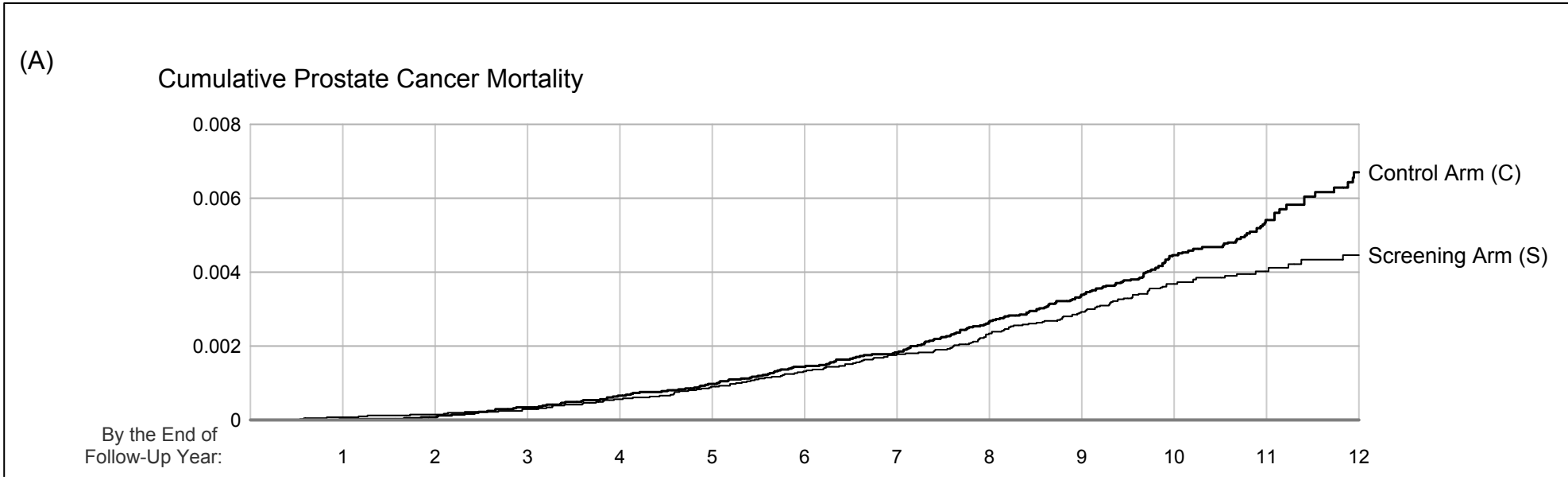
emphasis on time-specificity

- Year-specific* mortality rate ratios
- Moving averages* to reduce the statistical noise (deaths in moving 3-year intervals)
- Smooth curve for rate ratio function (data bins 0.2 y wide).

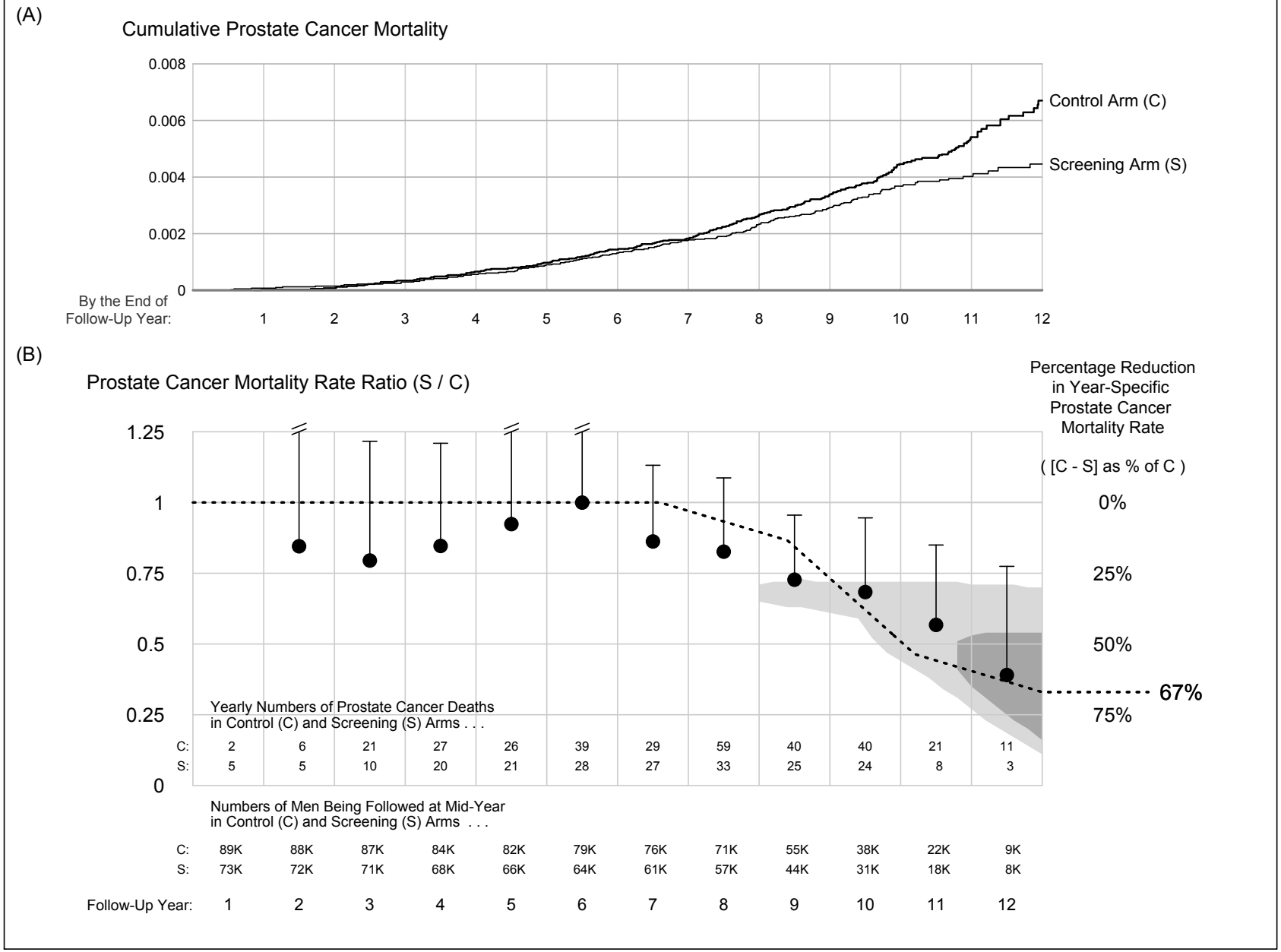
---

* cf. Miettinen et al. 2002

Year-specific prostate cancer mortality ratios



Year-specific prostate cancer mortality ratios



BREAST CANCER

IN EVERY INSTANCE: REDUCTION UNDER-ESTIMATED

See

Miettinen et al., *Lancet* 2002;

Hanley, *Epidemiologic Reviews* 2011.

LUNG CANCER

Mayo Lung Project (chest x-ray & sputum cytology)

- Enrollment: 1971-1976;
  negative on 'prevalence' screen;
  screening every 4 mo. for 6 years (vs., on enrollment, recommendation to receive annual chest x-ray & sputum cytology).
- JNCI 2000: "Lung Cancer Mortality in the Mayo Lung Project: Impact of Extended Follow-up"

  Would 24-year follow up "allow for a reduction in lung cancer mortality to be observed?"

- ALL lung cancer deaths, from those in year...
  - 1, before impact could become evident,
    to
  - 24, 18 years after last screen.

National Lung Screening Trial (NLST)

- Enrollment: August 2002 - March-2004
  3 annual screens: low-dose helical CT (vs. standard chest X-ray).
  Primary scientific goal:

  to determine whether three annual screenings with low-dose helical computerized tomography (LDCT) reduces [sic] mortality from lung cancer

- Press Releases, November 2010:

  Screening of people at high-risk for lung cancer with low dose CT significantly reduces lung cancer death: 20% fewer lung cancer deaths [ACR]

  An interim analysis of the study's primary endpoint, reported to the DSMB on October 20, 2010, revealed a deficit of lung cancer deaths in the LDCT arm, and the deficit exceeded that expected by chance, even allowing for the multiple analyses conducted during the course of the trial. Data presented at previous meetings of the DSMB did not meet the requirements for statistical significance with respect to the primary endpoint. [NCI(US)]

ACR Imaging Network: Press Release

Table 3: Interim Analysis of Primary Endpoint Reported on October 20, 2010

Trial Arm	Person years (py)	Lung cancer deaths	Lung cancer mortality per 100,000 py	Reduction in lung cancer mortality (%)	Value of test statistic	Efficacy boundary
LDCT	144,097.6	354	245.7	20.3	–3.21	–2.02
CXR	143,363.5	442	308.3			

"Deficit": 88

Timing of the ‘deficit’ of (442-354=) 88 deaths

? ? ?

Year:	1	2	3	4	5	6	7	8	ALL
Screens	↑	↑	↑						
? CXR arm:	10	38	65	75	82	90	60	22	442
?? LDCT arm:	10	36	59	59	56	63	50	21	354
?? deficit (no.):	0	-2	-6	-16	-26	-27	-10	-1	-88
?? deficit (%):	0%	5%	9%	21 %	32%	30%	17%	5%	20%
?? LDCT arm:	8	30	52	60	66	73	48	17	354
?? deficit (no.):	-2	-8	-13	-15	-16	-17	-12	-5	-88
?? deficit (%):	20%	21%	20%	20%	20%	19%	20%	23%	20%
?? LDCT arm:	?	?	?	?	?	?	?	?	354
?? deficit (no.):	-?	-?	-?	-?	-?	-?	-?	-?	-88
?? deficit (%):	? %	? %	? %	? %	? %	? %	? %	? %	20%

20% MORTALITY REDUCTION

A UNIVERSAL CONSTANT IN SCREENING TRIALS?

Reductions in 'event rates': 5 'prevention' studies

- HPV 6,11,16,18 infection:
  - Quadrivalent human papillomavirus (HPV) vaccine
- Paralytic or non-paralytic poliomyelitis:
  - Salk Vaccine
- HIV infection:
  - (Adult) Circumcision
- Death from ruptured abdominal aneurym:
  - Ultrasound screening
- Vascular events:
  - Statin treatment [elevated C-reactive protein at entry]

QUESTION: Shape of $\downarrow(t)$ function, i.e., % Reduction in Rate as function of follow-up time, if rates based on...

- all events up to that point in f-up time? (*1 'average' rate*) ?
- when in f-up time events occurred (*'time-specific'* rates) ?

(i) Percentage Reduction in AVERAGE Event Rate
(if data analyzed after indicated no. of events)

A Cervical intraepithelial neoplasia (HPV 6,11,16,18):
quadrivalent human papillomavirus (HPV) vaccine

B Paralytic or non-paralytoc poliomyelitis:
Salk Vaccine

C HIV:
Circumcision

D Death from ruptured abdominal aneurym:
Ultrasound screening

E Vascular events:
Statin treatment [elevated C-reactive protein at entry]

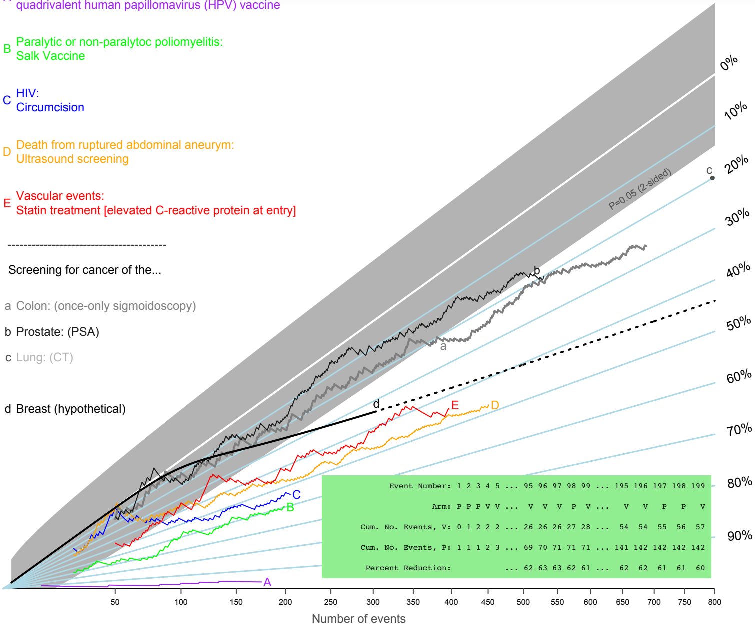
Screening for cancer of the...

a Colon: (once-only sigmoidoscopy)

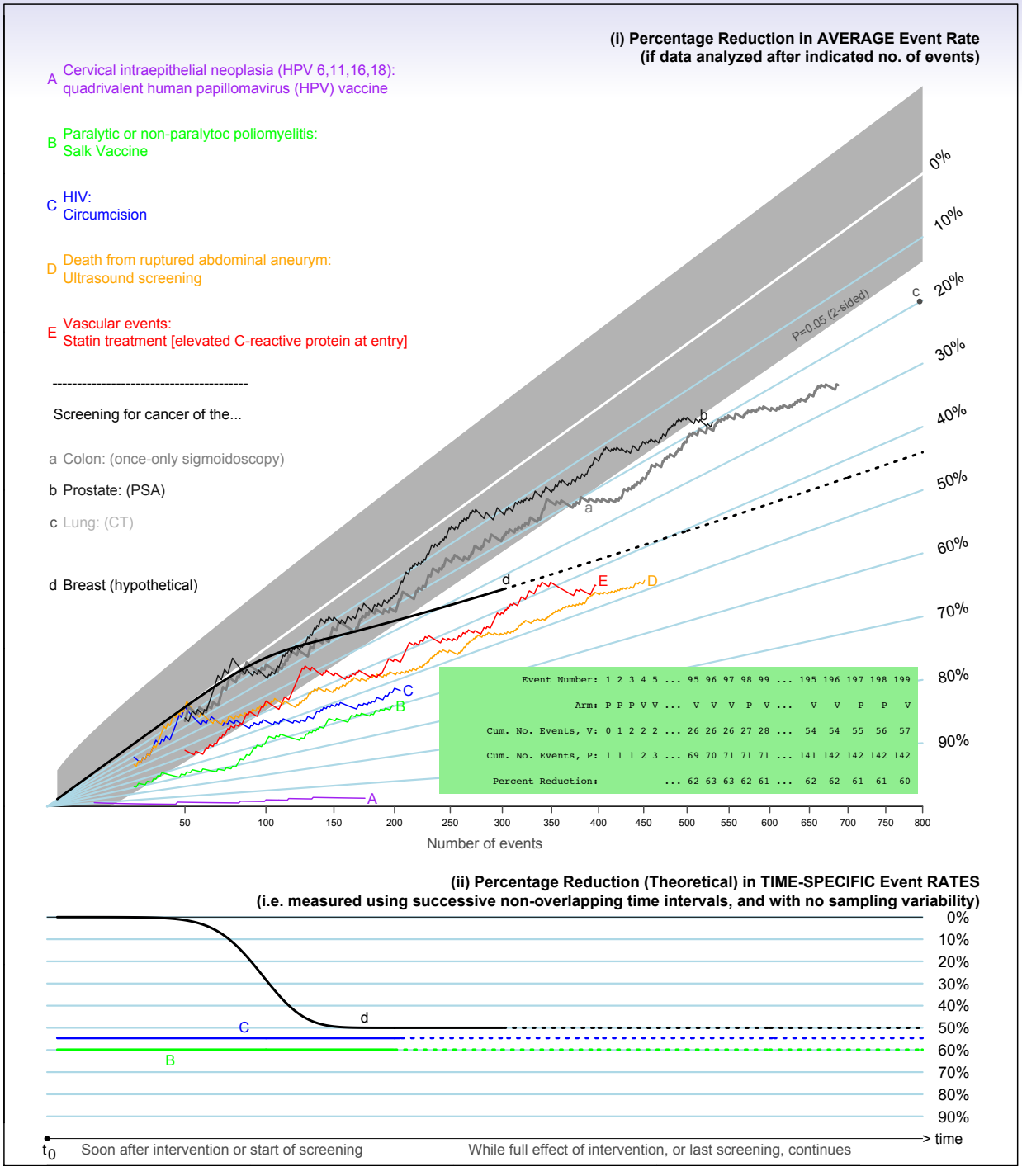
b Prostate: (PSA)

c Lung: (CT)

d Breast (hypothetical)



Event Number:	1	2	3	4	5	...	95	96	97	98	99	...	195	196	197	198	199
Arm:	P	P	P	V	V	...	V	V	V	P	V	...	V	V	P	P	V
Cum. No. Events, V:	0	1	2	2	2	...	26	26	26	27	28	...	54	54	55	56	57
Cum. No. Events, P:	1	1	1	2	3	...	69	70	71	71	71	...	141	142	142	142	142
Percent Reduction:						...	62	63	63	62	61	...	62	62	61	61	60



If **intervention continues over time to deflect the same % of events**, an estimate of the % reduction, based on the total number events in more (person)-time will be more precise

Mortality reductions from cancer screening manifest distally. Enrolling and following more people for short length of time yields a more precise UNDERestimate.

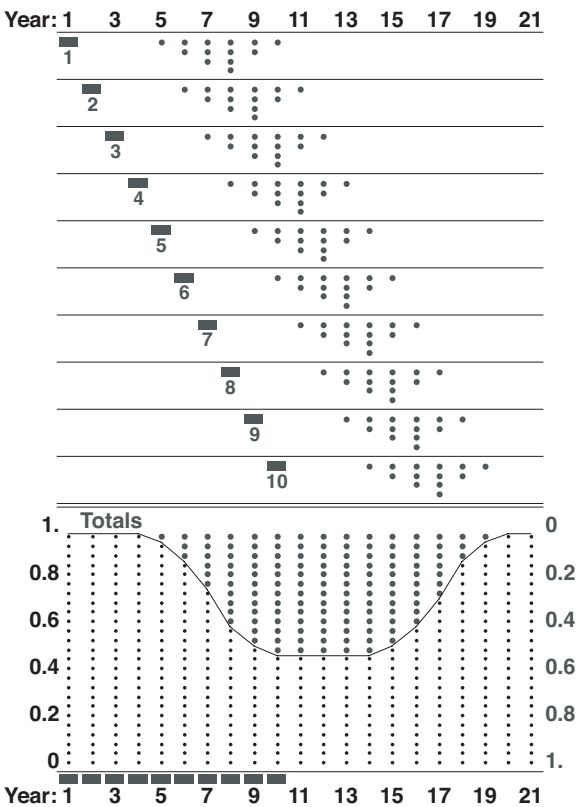
The seemingly-universal 20% reduction is an artifact of prevailing data-analysis methods and stopping rules.

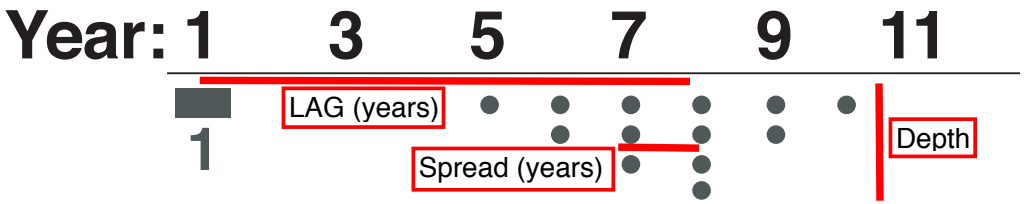
If use all data from time screening commences, the first % reduction which was statistically different from zero does not answer the question of interest to payers.

PLANS

Data and Methods, Parameters, their Use

- **Data**: completed RCTs of screening for prostate, breast, colon and lung ca; population-based screening programs.
- **3 Parameters** (‘*deliverables*’) and how they will be fitted:





y = years since screening commenced

- Rate ratio in Year *y*, Age *a* in Study *s* :

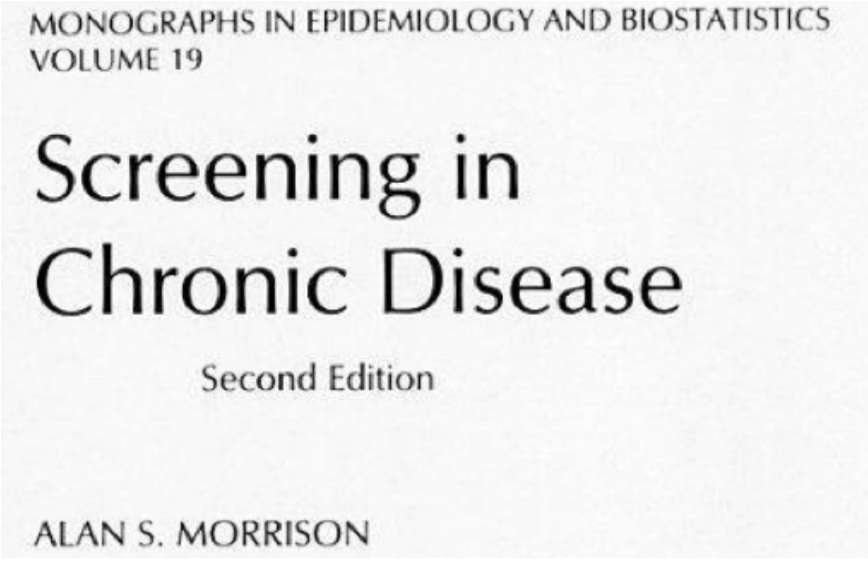
RateRatio(*y, a, s*) =

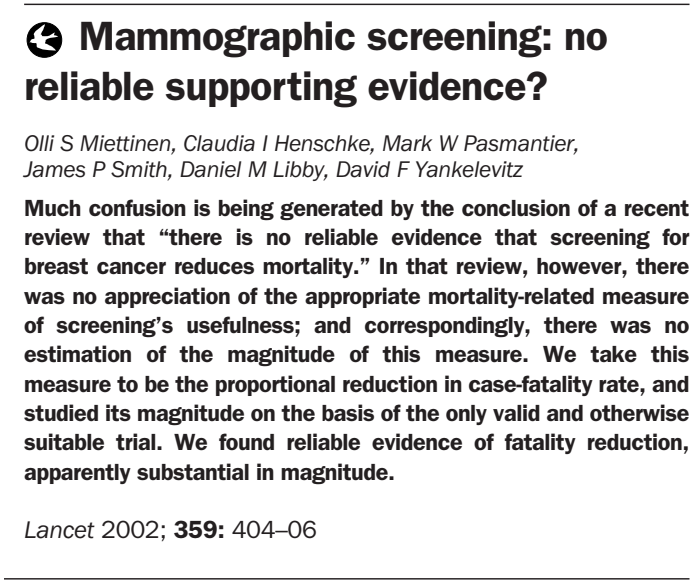
sum of reductions from all previous rounds of screening in study *s*

- Design matrix: 1 row per y-a-s ‘cell’
- $\frac{\text{No. deaths in screening arm}}{\text{No. deaths in 2 arms combined}}$ in each ‘cell’
- Fit by Max. Likelihood (binomial model)

- **USE**: project mort. reductions due to a screening regimen

Acknowledgments





Mammographic screening: no reliable supporting evidence?

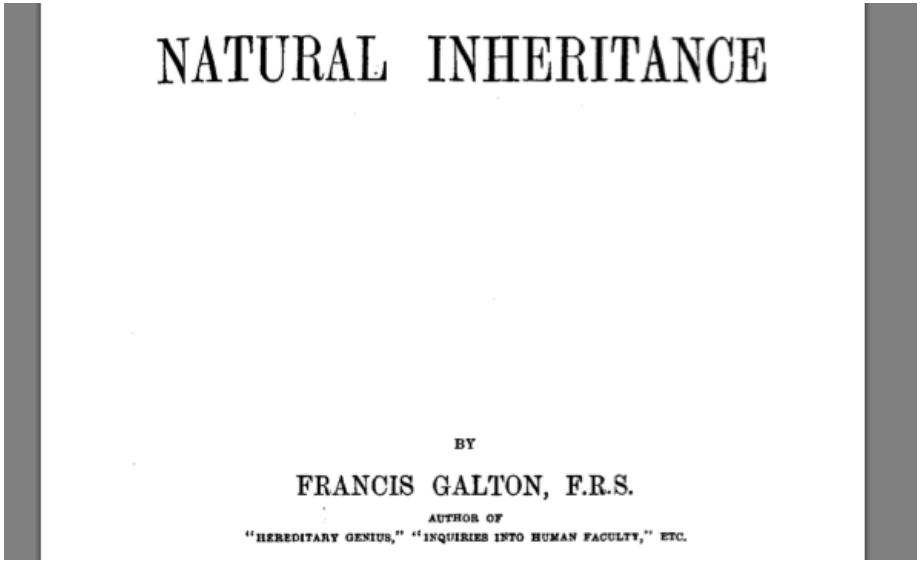
Olli S Miettinen, Claudia I Henschke, Mark W Pasmantier, James P Smith, Daniel M Libby, David F Yankelevitz

Much confusion is being generated by the conclusion of a recent review that "there is no reliable evidence that screening for breast cancer reduces mortality." In that review, however, there was no appreciation of the appropriate mortality-related measure of screening's usefulness; and correspondingly, there was no estimation of the magnitude of this measure. We take this measure to be the proportional reduction in case-fatality rate, and studied its magnitude on the basis of the only valid and otherwise suitable trial. We found reliable evidence of fatality reduction, apparently substantial in magnitude.

Lancet 2002; **359:** 404–06

Screening for breast cancer in women aged 40-49 years.
Montreal: CETS Report no. 22, 1993. 91p. Available at: http://www.aetmis.gouv.qc.ca/en/

J. Caro and M. McGregor



NATURAL INHERITANCE

BY

FRANCIS GALTON, F.R.S.

AUTHOR OF

"HEREDITARY GENIUS," "INQUIRIES INTO HUMAN FACULTY," ETC.

Why do statisticians commonly limit their inquiries to Averages?

F. Galton, Natural Inheritance, 1889.

“It is difficult to understand why statisticians commonly limit their inquiries to Averages, and do not revel in more comprehensive views.

Their souls seem as dull to the charm of variety as that of the native of one of our flat English counties, whose retrospect of Switzerland was that, if its mountains could be thrown into its lakes, two nuisances would be got rid of at once.”

Summary: my 3 points again

- With their blindness to the delay until the reductions in mortality are expressed, the prevailing design and data-analysis of cancer screening trials ***under-estimate*** the mortality reductions that ***would be produced by a sustained screening program***
- P-value-driven stopping rules exacerbate the underestimation
- We *might* be able to avoid such misleading numbers if we
  (i) recognize the issue, and avoid the standard RCT paradigm
  (ii) run trials with sufficient rounds of screening and sufficient follow-up
  (iii) spend major portion of career waiting to measure real reductions
  (iv) analyze the data using time-specificity
  (v) focus on the parameters that describe impact of 1 round of screening

FUNDING, CO-ORDINATES, DOWNLOADS

Natural Sciences and Engineering Research Council of Canada

Le Fonds québécois de la recherche sur la nature et les technologies

James.Hanley@McGill.CA

http://www.biostat.mcgill.ca/hanley

→ **r e p r i n t s / talks**

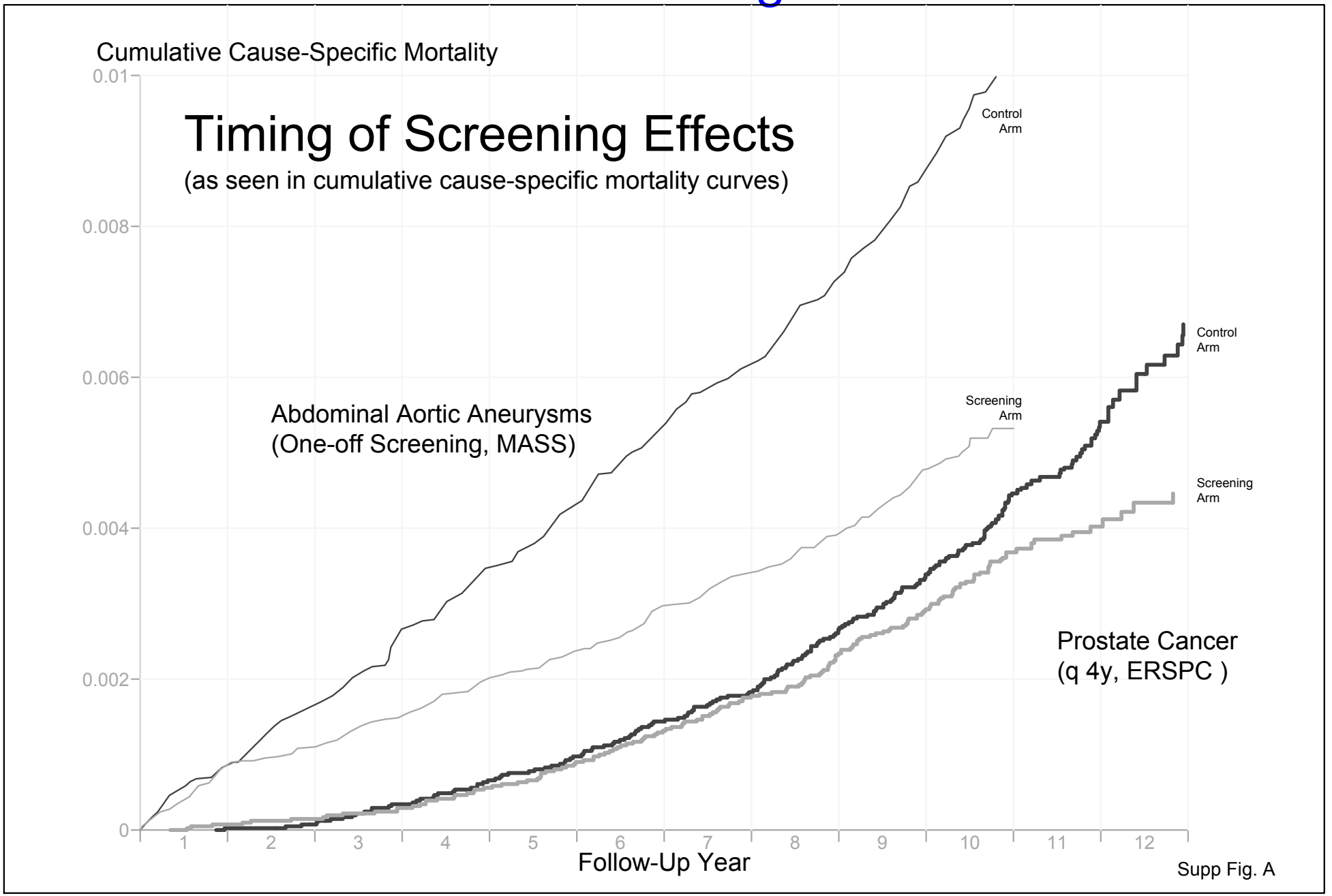
McGill **Biostatistics**
Biostatistique

http://www.mcgill.ca/epi-biostat-occh/grad/biostatistics/

Some References

1. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* 2009;360:1320-1328.
2. Hanley JA. Mortality reductions produced by sustained prostate cancer screening have been underestimated. Journal of Medical Screening. *J Medical Screening* 2010;17:147-151.
3. Hanley JA. Measuring Mortality reductions in cancer screening studies. *Epidemiologic Reviews* 2011. Advance Access published May 30, 2011.
4. Hanley JA. CANNeCTIN Clinical Trials Methodology Seminar Series. Videoconference April 9, 2010. Slides: http://www.cannectin.ca/ . Video: Archived Events, http://webcast.otn.ca/
5. Thompson SG, Ashton HA, Gao L, Scott RAP on behalf of the Multicentre Aneurysm Screening Study Group. Screening men for abdominal aortic aneurysm: 10 year mortality and cost effectiveness results from the randomised Multicentre Aneurysm Screening Study. *BMJ* 2009;338:b2307 doi:10.1136/bmj.b2307.
6. Hanley JA. Analysis of Mortality Data From Cancer Screening Studies: Looking in the Right Window. *Epidemiology* 2005; 16: 786-790.
7. Miettinen OS, Henschke CI, Pasmantier MW, et al. Mammographic screening: no reliable supporting evidence? *Lancet* 2002;359:404-406.
8. Miettinen OS, Henschke CI, Pasmantier MW, et al. Mammographic screening: no reliable supporting evidence? Available at: http://image.thelancet.com/extras/1093web.pdf. Accessed July 6, 2005.

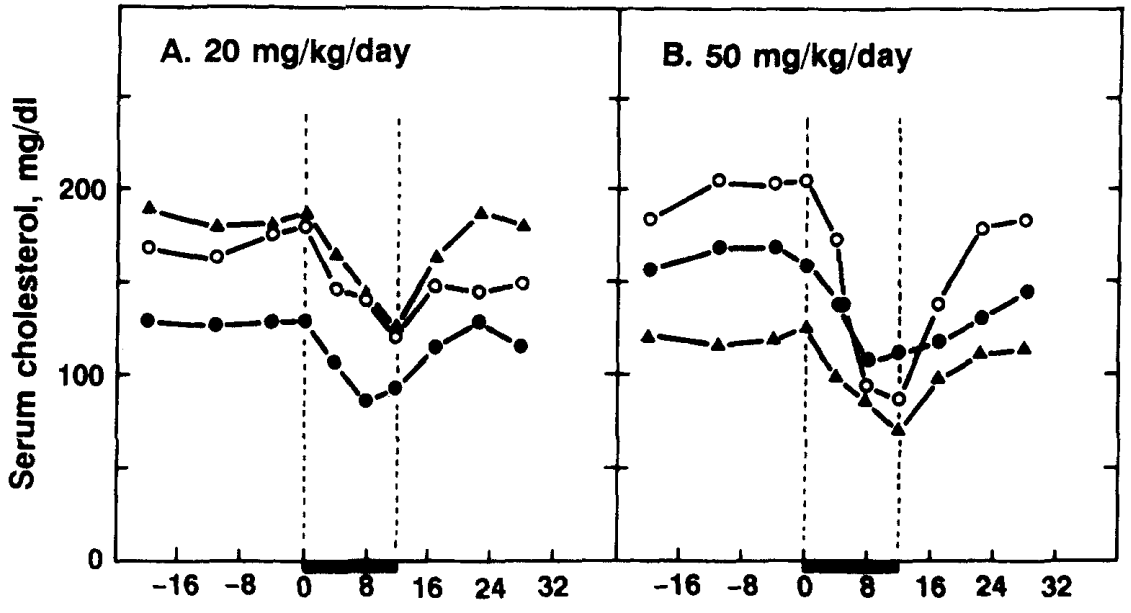
The loneliness of the long-distance trialist



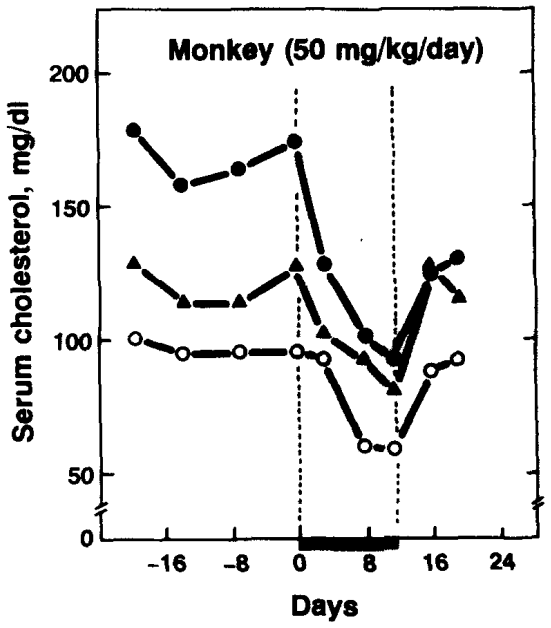
Supp Fig. A

Timing of cholesterol reductions produced by statins

3 dogs at 20 mg/kg/day; 3 at 50 mg/kg/day

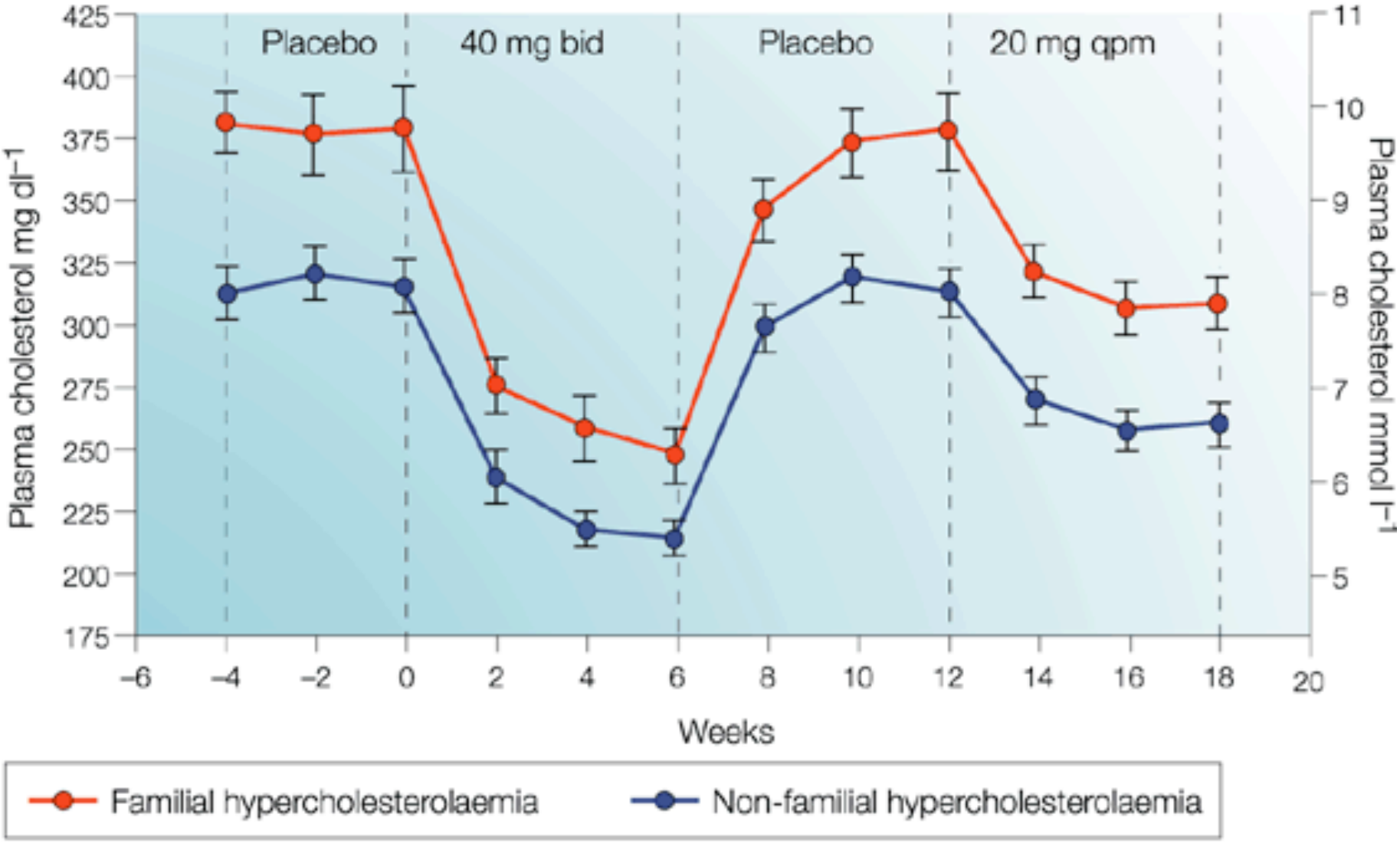


3 monkeys at 50



Timing of cholesterol reductions produced by statins

Humans



Norway - 'before-after' study

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 SEPTEMBER 23, 2010 VOL. 363 NO. 13

Effect of Screening Mammography on Breast-Cancer Mortality in Norway

Mette Kalager, M.D., Marvin Zelen, Ph.D., Frøydis Langmark, M.D., and Hans-Olov Adami, M.D., Ph.D.

Screening program was started in 1996 and expanded geographically during the subsequent 9 years.

Women between the ages of 50 and 69 years were offered screening mammography every 2 years.

Results & Conclusions

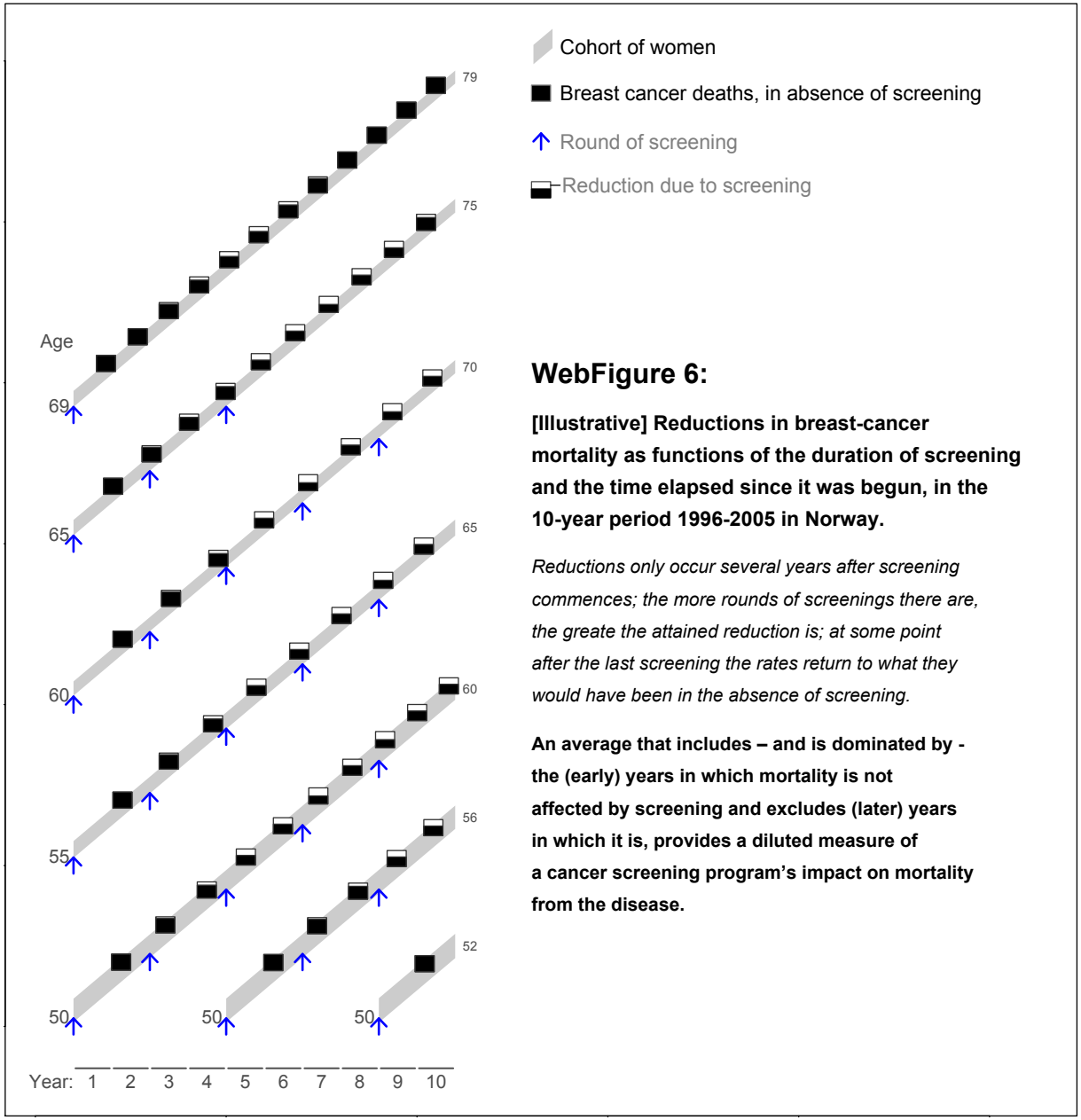
The rate of death was reduced by 7.2 deaths per 100,000 person-years in the screening group as compared with the historical screening group (rate ratio, 0.72; and by 4.8 deaths per 100,000 person-years in the nonscreening group as compared with the historical nonscreening group (rate ratio, 0.82; for a relative reduction in mortality of 10% in the screening group. Thus, the difference in the reduction in mortality between the current and historical groups that could be attributed to screening alone was 2.4 deaths per 100,000 person-years, or a third of the total reduction of 7.2 deaths. The availability of screening mammography was associated with a reduction in the rate of death from breast cancer, but the screening itself accounted for only about a third of the total reduction.

Time-insensitivity: not exclusive to RCT reports

Paraphrase of (refused) letter by JH to NEJM re 2010 analysis of data from Norway

Kalager Zelen
Langmark Adami.

Epidemiologic Reviews, 2011



WebFigure 6:

[Illustrative] Reductions in breast-cancer mortality as functions of the duration of screening and the time elapsed since it was begun, in the 10-year period 1996-2005 in Norway.

Reductions only occur several years after screening commences; the more rounds of screenings there are, the greate the attained reduction is; at some point after the last screening the rates return to what they would have been in the absence of screening.

An average that includes – and is dominated by - the (early) years in which mortality is not affected by screening and excludes (later) years in which it is, provides a diluted measure of a cancer screening program's impact on mortality from the disease.